Molecular Systems Biology 6; Article number 361; doi:10.1038/msb.2010.15
Citation: *Molecular Systems Biology* 6:361

www.molecularsystemsbiology.com

molecular systems biology

EDITORIAL

Reconstruction annotation jamborees: a community approach to systems biology

Molecular Systems Biology **6**; 361; published online 13 April 2010; doi:10.1038/msb.2010.15

Genome-scale metabolic network reconstructions represent biochemical, genetic, and genomic (BiGG) knowledge bases for a target organism (Reed *et al*, 2006). Thus, they correspond to two-dimensional genome annotations: that is, they contain all nodes and links that comprise a biochemical reaction network defined by the genome (Palsson, 2004). These reconstructions allow the conversion of biological knowledge into a mathematical format and subsequent computation of physiological properties. They therefore enable the formulation of a mechanistic genotype–phenotype relationship for metabolic functions in the target organism.

The metabolic network reconstruction process is now well established (Thiele and Palsson, 2010) and its workflows have recently been reviewed (Reed *et al*, 2006; Feist *et al*, 2009). The development of a consensus network reconstruction that is accepted and used by the research community necessitates a collective effort to formalize such networks that are specific to a target organism. This need has led to the concept of a 2D annotation (or a reconstruction) jamboree (Mo and Palsson, 2009), in analogy to the 1D genome annotation jamborees that lead to a community-driven genome annotation process. You may be interested in organizing a jamboree for your favorite target organism. What do you need to do?

Goals of a reconstruction jamboree

The goal of a network reconstruction jamboree is to reconcile and refine currently available BiGG knowledge about the target organism. If available, multiple existing metabolic network reconstructions made by individual research groups provide a great starting point. A jamboree should update, re-evaluate, refine and, later on, expand the network content. These goals are most efficiently achieved through a community approach that assembles experts from different areas.

Jamborees assist in fostering collaborations as well as informing the community about the properties, content, and capabilities of the consensus reconstruction to ensure its broad use for biological, biomedical, and biotechnological applications. It is important to establish standards and criteria that guide the jamboree teams. To date, 2D annotation jamborees have been carried out for three target organisms (*Saccharomyces cerevisiae* (Herrgard *et al*, 2008), *Salmonella typhimurium* LT2, and *Homo sapiens*).

Who should participate?

The jamboree team has to tackle many different issues and tasks to obtain a target organism-specific consensus reconstruction. Hence, it is important to invite experts in systems biology (for modeling); chemistry and metabolomics (for metabolite information); biochemistry, molecular and cell biology (for reaction and genetic information); and bioinformatics (for gene annotation and database structure).

Information that needs evaluation

At least three areas of metabolic reconstructions require currently detailed attention by the jamboree team, which include metabolites, metabolic reactions, and the gene–protein–reaction (GPR) associations. The information that needs to be associated with these areas has been described in detail in Thiele and Palsson (2010) and is summarized in Box 1.

Reconstruction versus model

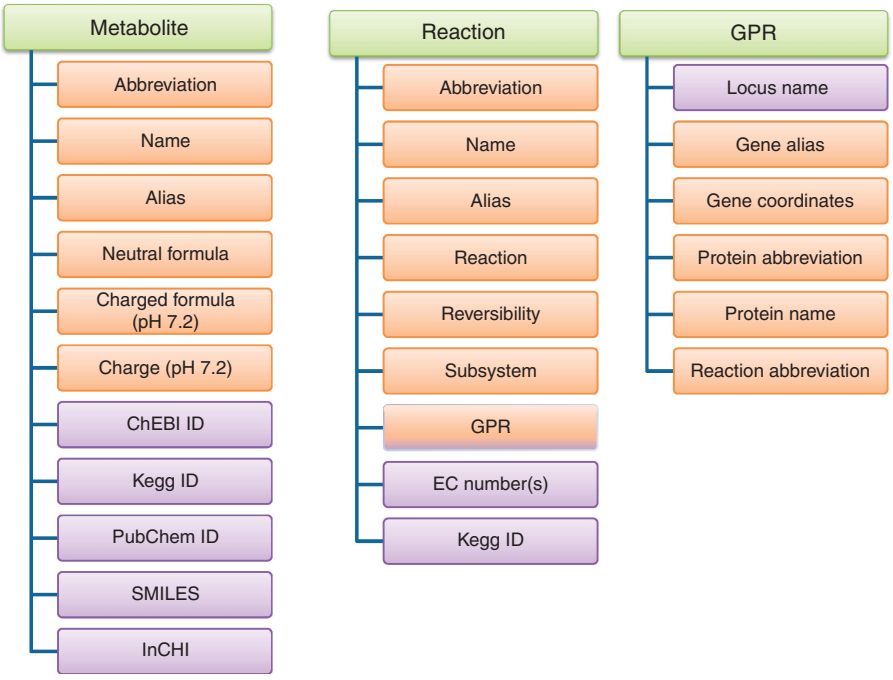
Using a confidence-scoring system, one can readily identify reactions with different levels of experimental support. With such a system in place, the consensus reconstruction can be readily converted into a mathematical model, but also allows for rapid elimination of low-confidence reactions if necessary (e.g., for high-throughput data mapping or visualization). Thus, the consensus reconstruction can cover all knowledge about the target organism, while it also highlights included uncertainties, which may be important for biological discovery projects. The use of a confidence scheme therefore allows satisfying both purposes of the consensus reconstruction.

Workflow for 'how-to do' a reconstruction jamboree

The experience with the first three reconstruction jamborees highlighted that reconstruction protocols and methods need to be standardized to facilitate an optimal outcome for consensus metabolic reconstructions. The same is true with the format and protocol of the reconstruction jamboree itself.

A current workflow for reconstruction jamborees is illustrated in Box 2. It consists of three phases: preparation, jamboree meeting, and wrap-up phase. While the jamboree

Box 1 Various data need to be associated with each metabolite and each reaction in the consensus reconstruction



Box 1 In orange are highlighted information that link to the reconstruction. The purple information links to other databases. *Metabolites*: The consensus reconstruction needs to have unique, standardized metabolite identifiers, e.g., ChEBI (Brooksbank *et al*, 2005), and database-independent metabolite representations, e.g., SMILES (Weininger, 1988). Chemical formula and charge have to be included to enable mass- and charge-balancing of network reactions. *Reactions*: Once metabolites are associated with unique identifiers, reactions can be easily recognized, making reaction identifiers superfluous. However, enzyme commission numbers (EC numbers) may be added as a global identifier. Reactions may use different substrates or cofactors, reaction stoichiometry and reaction directionality depending on the organism. For the reaction directionality, thermodynamic information need to be considered (e.g., as in Feist *et al*, 2007; Fleming *et al*, 2009). *Gene–protein–reaction (GPR) associations*: GPRs are included in most genome-scale, metabolic reconstructions and are encoded using Boolean expressions describing the gene product(s) responsible for catalysis of one or more network reactions. These expressions are perhaps most susceptible to discrepancies, as they rely mainly on interpretation of scientific literature, which is especially true for complex rules (e.g., protein complexes).

meeting (phase 2) involved the research community to tackle the reconciliation criteria described above, the preparation phase is typically executed by a small number of researchers ('jamboree council'). This first phase is, naturally, most important for the success of the jamboree and requires some time. The duration of the jamboree meeting will be directly dictated by the time availability of the participants and the amount of material to be evaluated, but may range from 2 to 5 days. Ideally, many of the issues are addressed during the jamboree meeting, and can be assembled and compiled in the wrap-up phase to form a first version of the consensus reconstruction.

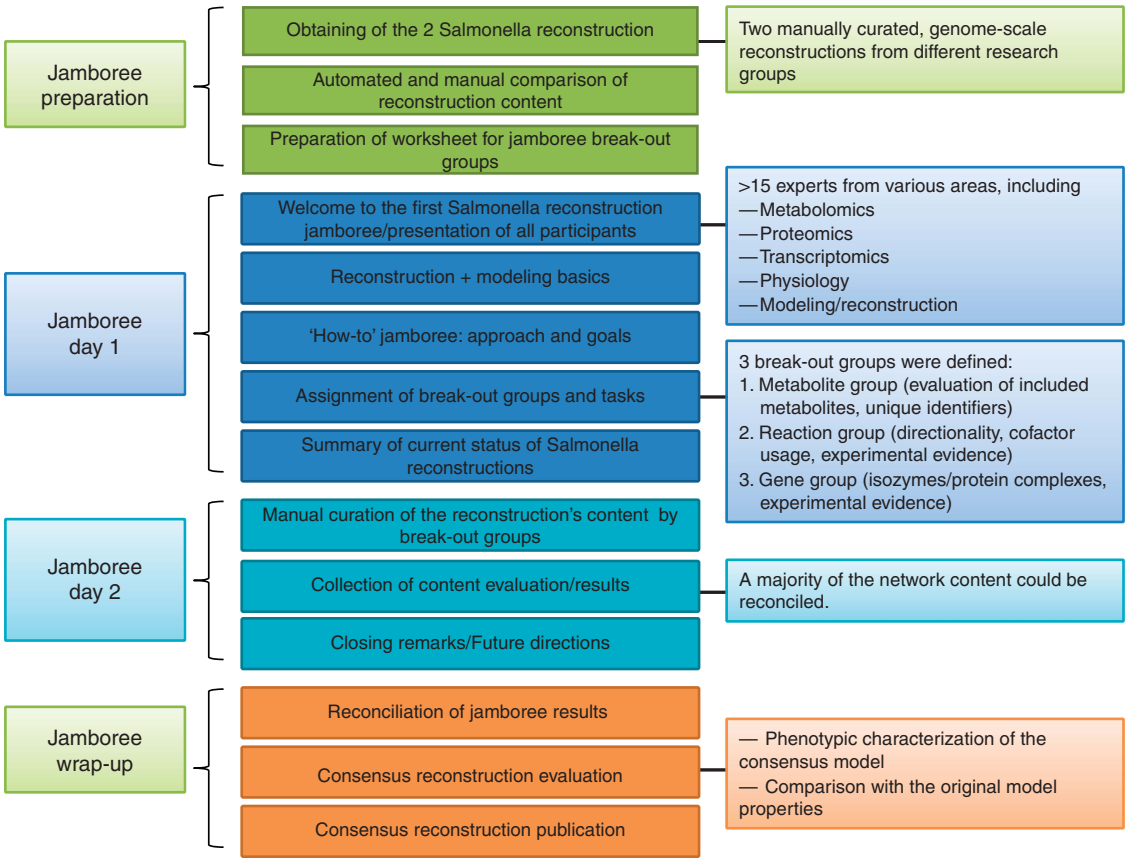
Evaluation criteria and standards during the jamboree meeting

The aforementioned goals need to be well structured and clear to all participants. Ideally, an information session about established reconstruction procedures as well as evaluation criteria (see above), followed by a Q&A session, is organized before the hands-on work. This will ensure the quality and value of the jamboree work. During the jamboree, the curators need to provide evidence for their decision (to keep/alter/reject a reaction or GPR association), in the form of publication references and notes. Although notes have the disadvantage that they are readable only by humans, they ensure tractability of the decisions and consensus reconstruction content.

Evaluation and decision criteria need to be established, e.g., how current knowledge is evaluated, which reactions/genes should be kept and based on which evidence. This issue becomes particularly important if contradicting results have been published in scientific literature. Since the consensus reconstruction reflects current knowledge, all results should be reported and connected with the consensus reconstruction (e.g., in the form of notes and confidence score). Based on this information, the 'jamboree council' will need to decide whether a reaction or gene is correctly included in the consensus reconstruction and document decision and evidence accordingly to make the decision tractable for other curators and users. As new information becomes available, this decision will need to be revisited.

Furthermore, experimental evidence that has been obtained not from the target organism but from related species needs to be highlighted by the curator (e.g., those that are important for human metabolic reconstruction or for less well-studied organisms). An appropriate vocabulary needs to be established.

Box 2 Workflow for the Salmonella reconstruction jamboree conducted on 5 and 6 September 2008 at the University of Iceland



Box 2 This workflow should serve as a template for organization of future metabolic reconstruction jamborees as it highlights the important steps and features of the consensus reconstruction. Phase 1: This phase should establish similarities and discrepancies between the metabolic reconstructions in terms of metabolites, reactions, and GPRs. This phase may require significant manual evaluation of the content but no reconciliation. The preparation phase should result in worksheets that state the problem that the jamboree team needs to address. Phase 2: During the jamboree meeting the participants will be divided into at least three groups based on preference and expertise (metabolite, reaction, and GPR group). A fourth group may be established for evaluation of reactions that have no evidence but may be needed for mathematical modeling. Each group will evaluate the material based on evidence given by the reconstruction and available resources (literature, databases, and annotations). Phase 3: The wrap-up phase will also include testing of the network functionality and comparison with the prediction capabilities of the initial reconstruction(s). Reconstruction dissemination will also be done in this phase. Follow-up meetings should be discussed to achieve the set goals and further refine the consensus reconstruction.

In closing

A 2D annotation jamboree provides a forum for bringing researchers together to build an organism-specific BiGG knowledge base, and for fostering ensuing collaboration and scientific communication. Ideally, a jamboree should be held regularly, e.g., every other year, depending on the community size around the target organism, availability of new data (e.g., biochemical, genetic, proteomic, metabolomic), and integration of additional cellular functions (e.g., signaling pathways, transcriptional regulation, etc.). For example, the second yeast reconstruction jamboree is currently planned (Pedro Mendes, personal communication). The continuous update will ensure that the consensus reconstruction will serve as a starting point for question- and condition-specific models, as well as that new experimental evidence, which may be derived from the reconciliation, is captured and incorporated. A well-crafted and executed reconstruction jamboree should accelerate the understanding of the systems biology of the target organism, as well as provide the platform for targeted experimental investigation for biological discovery, understanding, and synthesis.

Acknowledgements

We are thankful to Markus Herrgard, Monica L Mo, Ronan MT Fleming, Neil Swainston, Dirk Bumann, Pedro Mendes, and Douglas Kell for their valuable discussions. This work was supported by ERC grant no. 232816.

Conflict of Interest

The authors declare that they have no conflict of interest.

Ines Thiele[1,2] and Bernhard Ø Palsson[3]
[1]Center for Systems Biology, University of Iceland, Reykjavik, Iceland,
[2]Faculty of Industrial Engineering, Mechanical Engineering & Computer Science, University of Iceland, Reykjavik, Iceland and
[3]Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA

References

Brooksbank C, Cameron G, Thornton J (2005) The European Bioinformatics Institute's data resources: towards systems biology. *Nucleic Acids Res* **33:** D46–D53

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO (2007) A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* **3:** 121

Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO (2009) Reconstruction of biochemical networks in microorganisms. *Nat Rev* **7:** 129–143

Fleming RM, Thiele I, Nasheuer HP (2009) Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to Escherichia coli. *Biophys Chem* **145:** 47–56

Herrgard MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, Bluthgen N, Borger S, Costenoble R, Heinemann M, Hucka M, Le Novere N, Li P, Liebermeister W, Mo ML, Oliveira AP, Petranovic D, Pettifer S, Simeonidis E, Smallbone K *et al* (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol* **26:** 1155–1160

Mo ML, Palsson BO (2009) Understanding human metabolic physiology: a genome-to-systems approach. *Trends Biotechnol* **27:** 37–44

Palsson BO (2004) Two-dimensional annotation of genomes. *Nat Biotechnol* **22:** 1218–1219

Reed JL, Famili I, Thiele I, Palsson BO (2006) Towards multidimensional genome annotation. *Nat Rev* **7:** 130–141

Thiele I, Palsson BO (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protocols* **5:** 93–121

Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inform Comput Sci* **28:** 31–36